

Statistik für Informationsmanager

Klausur vom 4. August 2004

Name:

Matrikelnummer:

Erlaubte Hilfsmittel: Außer Schreibwerkzeug diesmal auch Taschenrechner; Antworten bitte im Anschluss an die Fragen, Rückseiten können für Antworten und Entwürfe genutzt werden; Entwürfe bitte anschließend durchstreichen!

Mit der elektronischen Veröffentlichung meiner Klausurnote zusammen mit meiner Matrikelnummer bin ich einverstanden:

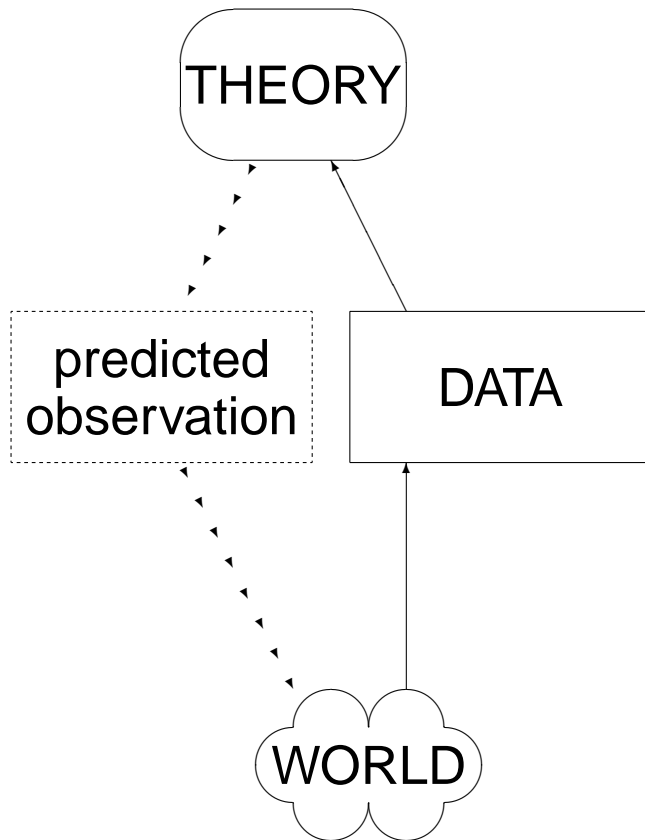
Ja Unterschrift
 Nein (keine Unterschrift heißt nein)

Für die Teilaufgaben wird es folgende Punktzahlen geben:

Aufgabe	Maximal erreichbar	Sie haben erreicht
1	$2 \cdot (1 + 2) = 6$	
2	$8 \cdot (3/8 + 3/8) = 6$	
3	$3 \cdot (1 + 1) = 6$	
4.1	2	
4.2	3	
4.3	5	
5	$3 \cdot 2 + 4 = 10$	
6	$2 \cdot 1 + 3 + 2 \cdot 2 = 9$	
Summe	57	

1. Grundlagen

Erinnern Sie sich an die mehrfach gezeigte Abbildung;



Zeichnen Sie ein, wohin

- a. Zeitreihenanalyse (1P)
- b. Clusteranalyse (1P)

hingehören! Geben Sie jeweils eine kurze Begründung!

- a. Zeitreihenanalyse gehört dorthin, weil ... (2P)

- b. Clusteranalyse gehört dorthin, weil ... (2P)

2. Skalenniveaus

Geben Sie jeweils für die folgenden Variablen an, ob sie nominal, ordinal, intervall- oder ratioskaliert sind (jeweils 3/8 P) und geben Sie jeweils eine kurze Begründung (jeweils weitere 3/8 P). (Eine gute Begründung für eine unserer Ansicht nach falsche Einordnung gibt eventuell ebenfalls Punkte, keine Begründung für eine unserer Ansicht nach richtige Einordnung gibt einen Punktabzug!)

	Nominal	Ordinal	Intervall	Ratio	Begründung
ECTS-Punkte für ein Modul in einem Studiengang					
Alkoholgehalt eines Getränks					
Matrikelnummer					
Zinssatz					
Datum					
Torverhältnis in einer Bundesliga-Tabelle					
Dauer einer Bahnfahrt					
Höhe eines Ortes über dem Meeresspiegel					

3. Univariate Statistik

Berechnen Sie (jeweils falls sinnvoll!) den Modus, den Median und den Mittelwert für folgende Variablen (richtige Berechnung[en] jeweils 1P, richtige Begründung jeweils noch 1P, wenn bei einer Teilaufgabe richtigerweise alle drei Werte berechnet worden sind, bedarf es keiner Begründung, und es gibt trotzdem 2P):

1. Preis für eine leere DVD unterschiedlicher Formate und unterschiedlicher Hersteller:

1.80	2.00	2.40	5.50	8.90	1.60	1.90	2.20	3.30	1.60
1.80	1.90	2.20	2.40	1.80	2.40	1.80	2.40	3.20	3.80
Mittelwert:									
Median									
Modus									
gegebenenfalls Begründung für nicht berechnete Parameter:									

2. Zufriedenheit mit der Gesundheit auf einer Skala von 0 („ganz unzufrieden“) bis 10 („vollauf zufrieden“) — Angaben der ersten 20 Befragten aus dem Sozioökonomischen Panel 2003

8	9	9	6	4	4	8	3	7	6
9	10	8	6	7	9	10	2	3	7
Mittelwert:									
Median									
Modus									
gegebenenfalls Begründung für nicht berechnete Parameter:									

3. Religionszugehörigkeit von 20 Befragten aus dem GSOEP 2003:

Ev	Kath	Kath	Ev	Sonst. Chr.	Sonst Rel.	Ohne	Ohne	Ev	Kath
Ohne	Ohne	Kath	Ev	Ohne	Ev	Kath	Ev	Kath	Ev
Mittelwert:									
Median									
Modus									
gegebenenfalls Begründung für nicht berechnete Parameter:									

5. Multivariate Statistik / Multiple Regression

In GSOEP 2003 wurde u.a. nach der bevorzugten wöchentlichen Arbeitszeit gefragt. Außerdem sollten sich alle Befragten zu ihrem Freizeitverhalten äußern (Achtung: die Zifferncodes bedeuten: 1 „täglich“, 2 „mindestens einmal in der Woche“, ..., 5 „nie“!). Die folgende Tabelle ist nur aus den Angaben der vollzeiterwerbstätigen Befragten berechnet worden und listet die Koeffizienten einer multiplen Regression auf 6 der 17 Freizeitvariablen auf (die Hinzunahme der anderen 11 hätte das Regressionsergebnis — $R^2 = 0.018$, $R = 0.137$ — nicht signifikant verbessert).

Koeffizienten

	Nicht standardisierte Koeffizienten B	Standardisierte Koeffizienten Beta	Signifikanz
(Konstante)	40.0501 h		<0.0005
Auto reparieren	-0.8208 h	-0.116	<0.0005
Klassische Konzerte, Theater etc.	+0.5270 h	+0.046	<0.0005
Reparaturen im Haus, Gartenarbeiten	+0.1945 h	+0.032	0.013
Kommunalpolitik	-0.3671 h	-0.026	0.030
Künstlerische und musikalische Aktivitäten	+0.2257 h	+0.032	0.013
Private Internetnutzung	-0.1326 h	-0.029	0.018

Abhängige Variable: gewünschte Wochenarbeitszeit

Was bedeuten die beiden Zahlen **40.0501 h** und **-0.8208 h** in den ersten beiden Zeilen der Tabelle? (2P).

Berechnen Sie die zu erwartende gewünschte Wochenarbeitszeit eines/r Befragten, der/die alle 6 Freizeitaktivitäten „täglich“ betreibt (d.h. jeweils „1“ angegeben hat) im Vergleich zu jemand, der/die alle 6 Freizeitaktivitäten „nie“ betreibt (d.h. jeweils „5“ angegeben hat). (2 P).

Nennen Sie die vier Freizeitaktivitäten, die sich nach diesem Analyseergebnis am stärksten auf das Nettoeinkommen auswirken, in der richtigen Reihenfolge. Beachten Sie dabei den Unterschied zwischen standardisierten und unstandardisierten Koeffizienten! (2 P)

Drücken Sie in einem ganzen Satz aus, was die Zahl **0.018** (am Ende der letzten Spalte) bzw. ihr Kehrwert 55 bedeutet! (3 P)

6. Multivariate Statistik / Faktorenanalyse

Ebenfalls im Sozioökonomischen Panel von 2003 wurde gefragt, wie zufrieden die Befragten mit verschiedenen Aspekten des Lebens waren. Die Zufriedenheit konnte auf einer Skala von 0 („völlig unzufrieden“) bis 10 („völlig zufrieden“) eingestuft werden.

Aus diesen 1 Fragen ließ sich die nachstehende Faktorladungsmatrix gewinnen. Koeffizienten kleiner als 0.2 sind zur Erleichterung des Auffindens einer Einfachstruktur weggelassen, außerdem sind die Variablen nach ihren höchsten Faktorladungen sortiert. Interpretieren Sie die beiden Faktoren, indem Sie ihnen möglichst prägnante Namen geben. Begründen Sie Ihre Namensgebung.

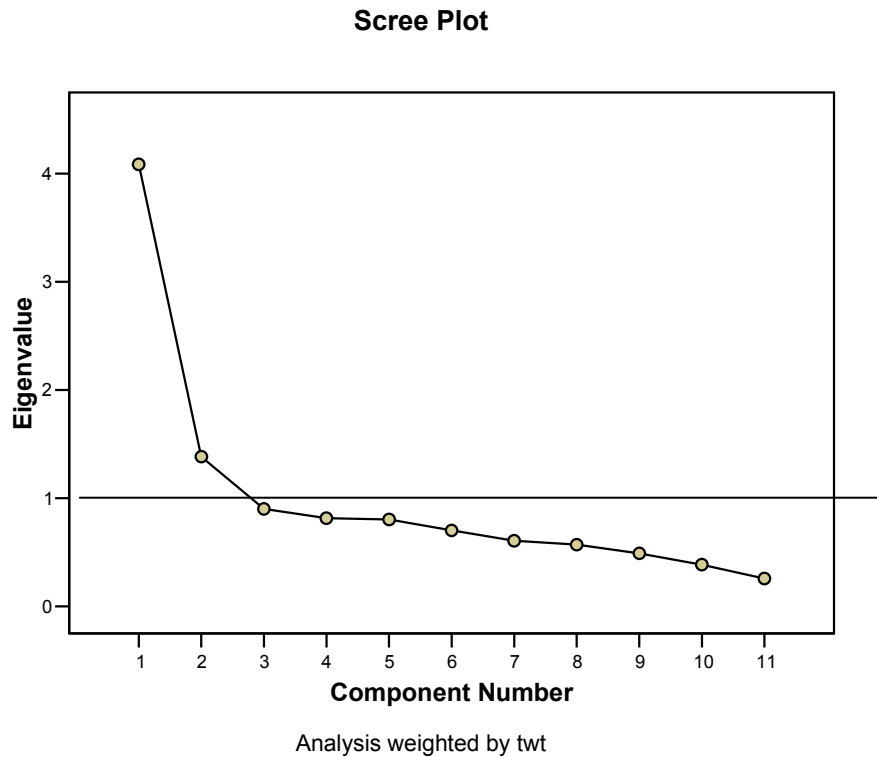
Zufriedenheit mit ...	Faktoren	
	1	2
Lebensstandard	.819	.202
Haushaltseinkommen	.764	
Wohnverhältnissen	.668	
Arbeit	.656	
Gesundheit	.609	
Häuslicher Arbeit	.587	
Umfang der Freizeit	.537	.220
Öffentlichen Diensten		.864
Öffentlichem Nahverkehr		.858
Kinderbetreuung	.332	.444
Umwelt	.424	.433

Faktor 1 beschreibt die latente Eigenschaft ... (1 P)

Faktor 2 ... (1 P)

Wie beurteilen Sie die Einfachstruktur? Widersprechen die mehrfachen hohen Faktorladungen in den letzten beiden Zeilen Ihrer Interpretation oder stützen sie sie? Warum? (3 P)

Hier finden Sie den Scree-Plot zu der vorstehenden Auswertung:



Halten Sie danach die Extraktion von zwei Faktoren noch für gerechtfertigt? (Ja / Nein / Kommt darauf an)
Wenn ja: warum? (2 P)

Wenn nein: warum nicht? Wie viele hätte man statt dessen extrahieren sollen? (2 P)

(Antworten auf beide Unterfragen werden gewertet! Wenn Sie dezidiert der Auffassung sind, dass nicht drei, sondern zwei oder vier oder fünf oder sechs oder sieben oder acht ... Faktoren hätten extrahiert werden sollen, müssen Sie auf „wenn nein: warum nicht?“ besonders ausführlich antworten; ebenso müssen Sie, wenn Sie dezidiert der Auffassung sind, dass drei Faktoren die richtige Wahl waren, auf „wenn ja, warum?“ besonders ausführlich antworten.)